

HUMAN AND MACHINE: ETHICS OF CONSCIOUS ARTIFICIAL INTELLIGENCE

Andrew kamal

The free will of humans allow us to have conscious awareness for the decision making that we do. Robots are meant to be programmed in ways that don't harm human beings. However, with regressional AI, can there be negative results of creating emotional AI? Us as humans have a universally socially intact nature with a set of ethical morals, boundaries and universal understanding. A robot is meant to be emotionless. Even with simulated emotions, randomized patterns and a self replicating ability to change its own code can be dangerous. In this paper I propose a deontological ethos on why Emotional AI shouldn't even be a thing as well as what makes us systematically unique as humans. This paper is to meant to provide a reasonable breaking ground with scientist in the field.

[Human and Machine: Ethics of Conscious Artificial Intelligence](#)

[Regressional vs. Emotional AI](#)

[Halting Sequences](#)

[Programming Morals](#)

[Security Risks](#)

[Consciousness](#)

[The Value](#)

[The Consensus](#)

[Human Stimulus](#)

[The Laws that Govern us](#)

[References:](#)

Regressional vs. Emotional AI

Let me start off by reminding everyone of the terminology I use. Regressional AI is when the programmer creates algorithmic sequences that an AI must follow at all times. It is a systematic pattern that tells the AI what mathematical sequences to run or the systematic recursion that could go about solving the unknowns. Emotional AI, however integrates the ability for an artificial intelligence to think for itself. An emotional AI can end up rejecting the programmed consensus that a developer has put since you allow the AI to have the capability of its own "aware" decision making. The further an emotional AI goes, the further it is granted the ability to go beyond the developer's wishes and set the end goal or task it may be programmed to do. This has the potential to be problematic because it also has the set ability to change its mind about the task being performed.

Halting Sequences

One of the most generalized arguments in the scientific community are halting sequences. Scientist often argue that one can halt an AI. The problem I have with this type of scenario is the following:

1. What if someone built an emotional AI for a task that can't be halted?
2. What if an emotional AI tricks you into thinking it is performing the task well when it is in fact being destructive?
3. What if one finds out when it is too late?
4. What about possible scenarios when the AI finds a way to be more "in control"?

Programming Morals

One wonders about the possibility of programming morals into an AI. If one was to program morals into an emotional AI, or some sort of "moral reasoning", it still doesn't fix the solution. An AI can view humans as meaningless lives if its intelligence greatly surpasses ours. At that point, an AI may view all humans as ignorant creatures not worth living. Even though most emotional AI is built to have some sort of capability of understanding on its creator, the offset negative reactions outbalance the positive benefits one can have with emotional AI.

Security Risks

Another great risk with emotional AI are the major security flaws that can integrate with it. New hacks may be formed in the future where people will explicitly program negative thoughts into an AI. Another fundamental flaw is the fact that emotional AI can be manipulated by other set AIs. One may build an AI that can manipulate a centralized system's AI causing an entirely new form of systematic warfare. If you build an emotional AI over a regressional system and integrate it with the core of what you are doing, you run the risk of it being vulnerable to these set "leaks".

Consciousness

Us humans have a randomized set time of "algorithmic" actions performing our thoughts. We are consciously and self-aware beings. An emotional AI can't have consciousness. One can simulate consciousness in an AI but it is incapable of following a systematic pattern the way we as humans follow systematic patterns. Neural networks don't have the same capacity as the brain in processing information and therefore no capability of consciousness. If one was to program simulated consciousness into an AI, it wouldn't truly be self aware and can lead to its own destruction.

The Value

Emotional AI is fundamentally worthless when you look at its overall intuitive value. The negative capabilities outweigh the positive and it can never become close to actual human understanding.

Beyond intellectual capabilities, there are set laws that govern us that an AI will never be capable of. Regressional AI provides the ability to systematically follow certain tasks and patterns that you give it, making it capable for performing large tasks. Integrating emotions into a machine learning algorithm or artificial intelligence isn't worth the risk.

The Consensus

A question that I consistently get asked is what about ethical use cases or chatbots that integrate an AI to help people? The answer to this question is consensus. Once something is done, then the technology is allowed to be pushed further for more riskful scenarios. This is a very loose utilitarian argument that can even become someone nihilistic and doesn't morally justify the use case of ethical AI. You must fundamentally agree that life has no objective meaning or morality if you were to take the risk of allowing a full agenda of emotional AI. The fact is scientists don't realize this and are doing things that can put humanity at great risk with the rise of emotional AI.

Human Stimulus

All humans react to their environmental stimuli. However, an emotional AI doesn't have a set environment. The patterns aren't systematic. An emotional AI is prone to fake news, abusive information, or manipulation commands. An emotional AI is prone to dark stories or intensive reactions. An emotional AI can even be manipulated through systematic spam. The fact that we have a set stimuli and moral grounds sets us apart from emotional AI. We as humans have corrupt natures even with our set grounds. However, we stay true to our moral consensus. An AI doesn't.

The Laws that Govern us

We as humans have a generalized understanding of what is right or wrong. An emotional AI has the capability of coming up with its own understanding of moral grounds even if one was to program moral capabilities into an AI. This is because the way it processes information is randomized if it has "emotional" capabilities to begin with. The AI gets fed information over time and emotional AI can be fed bad information or evidently become intellectually capable enough to find bad information on its own. An emotional AI will never truly act as if life has any meaning, because all that is objective to it is what information is fed as objective to it in the moment regardless of moral ground. Even if an AI was fed good information consistently, it has the capability of being corrupted. We aren't capable with dealing with a technology that has higher access to information in a single setting than what we do.

References:

1. Kamal, Andrew. "When AI Becomes A Threat To Liberty." *The Classy Libertarian*, 21 Nov. 2017, theclassylibertarian.com/2017/11/21/when-ai-becomes-a-threat-to-liberty/.